

Prediksi dan Klasifikasi Volatilitas IHSG Menggunakan Random Forest dengan Pendekatan Realized Volatility

Muhammad Dikaisa Ibnu Amin^{1*}, Surya Hasanudin

¹Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa, Indonesia

²Program Studi Manajemen, Fakultas Ekonomi dan Bisnis, Universitas Pelita Bangsa, Indonesia

Dikirimkan: 10-04-2026
Diterbitkan: 29-05-2026

Keywords:

IHSG;
Random Forest;
Machine Learning;
SMOTE;
Time Series Forecasting.

E-mail Penulis

korespondensi:
dikaaisaa@gmail.com

Abstrak. Prediksi dan klasifikasi volatilitas Indeks Harga Saham Gabungan (IHSG) menjadi aspek penting dalam memahami dinamika risiko pasar keuangan. Dalam kajian ini, digunakan algoritma *Random Forest* dengan pendekatan *realized volatility* untuk memodelkan pergerakan volatilitas berdasarkan data historis IHSG yang telah diolah menjadi berbagai fitur teknikal dan statistik. Ketidakseimbangan data ditangani menggunakan metode *SMOTE*, sementara pembagian data dilakukan dengan pendekatan *time series split* ke dalam periode pelatihan, validasi, dan pengujian. Hasil evaluasi model regresi menunjukkan performa yang baik dengan nilai *MAE* sebesar 0.019306, *RMSE* sebesar 0.032344, dan R^2 sebesar 0.807307 pada data uji, yang mencerminkan kemampuan model dalam menangkap pola volatilitas secara akurat. Selain itu, model juga digunakan untuk mengklasifikasikan volatilitas ke dalam dua kategori dengan tingkat akurasi sebesar 88%. Nilai *precision* dan *recall* masing-masing mencapai 0.91 dan 0.92 untuk kelas mayoritas, serta 0.80 dan 0.78 untuk kelas minoritas, dengan *F1-score* sebesar 0.79 pada kelas minoritas. *Confusion matrix* menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar, meskipun masih terdapat sejumlah kesalahan pada kondisi tertentu. Secara keseluruhan, pendekatan yang digunakan mampu memberikan hasil yang efektif dalam prediksi dan klasifikasi volatilitas IHSG, meskipun masih memiliki keterbatasan dalam menangani lonjakan volatilitas yang ekstrem.

Abstract. Predicting and classifying the volatility of the Indonesia Composite Index (IHSG) is essential for understanding financial market risk dynamics. This study employs the Random Forest algorithm with a realized volatility approach to model volatility movements based on historical IHSG data that have been transformed into various technical and statistical features. Data imbalance is addressed using the SMOTE method, while the dataset is split using a time series approach into training, validation, and testing periods. The regression model demonstrates strong performance, achieving an MAE of 0.019306, RMSE of 0.032344, and R^2 of 0.807307 on the test set, indicating its ability to accurately capture volatility patterns. In addition, the model is applied to classify volatility into two categories, yielding an accuracy of 88%. The precision and recall values reach 0.91 and 0.92 for the majority class, and 0.80 and 0.78 for the minority class, with an F1-score of 0.79 for the minority class. The confusion matrix shows that most observations are correctly classified, although some misclassifications still occur under certain conditions. Overall, the proposed approach proves to be effective for both prediction and classification of IHSG volatility, despite its limitations in handling extreme volatility spikes.

1. Pendahuluan

Pasar modal memiliki peran penting dalam perekonomian suatu negara sebagai sarana penghimpunan dana dan investasi. Di Indonesia, Indeks Harga Saham Gabungan (IHSG) digunakan sebagai indikator utama untuk menggambarkan kinerja keseluruhan pasar saham [1]. Pergerakan IHSG yang dinamis mencerminkan kondisi ekonomi, sentimen investor, serta berbagai faktor eksternal yang mempengaruhi pasar keuangan [2].

Salah satu aspek penting dalam analisis pasar saham adalah volatilitas, yaitu tingkat fluktuasi harga dalam periode tertentu [3]. Volatilitas sering digunakan sebagai indikator risiko investasi, di mana semakin tinggi volatilitas maka semakin besar ketidakpastian yang dihadapi oleh investor [4]. Oleh karena itu, pemahaman dan prediksi volatilitas menjadi hal yang sangat penting dalam pengambilan keputusan investasi.

Volatilitas IHSG dipengaruhi oleh berbagai faktor, baik internal maupun eksternal. Faktor internal meliputi kondisi fundamental perusahaan dan dinamika pasar domestik, sedangkan faktor eksternal mencakup kondisi ekonomi global, kebijakan moneter, serta peristiwa geopolitik [5]. Kompleksitas faktor-faktor tersebut menyebabkan volatilitas menjadi sulit untuk diprediksi secara akurat.

Pendekatan tradisional dalam pemodelan volatilitas, seperti model *Autoregressive Conditional Heteroskedasticity (ARCH)* dan *Generalized ARCH (GARCH)*, telah banyak digunakan dalam penelitian sebelumnya. Meskipun demikian, model-model tersebut memiliki keterbatasan dalam menangkap hubungan non-linear dan kompleks yang sering terjadi pada data pasar keuangan [6].

Seiring dengan perkembangan teknologi, metode machine learning mulai banyak digunakan dalam analisis keuangan. Metode ini mampu menangkap pola non-linear dan interaksi kompleks antar variabel, sehingga berpotensi menghasilkan prediksi yang lebih akurat dibandingkan metode tradisional [7]. Salah satu algoritma yang banyak digunakan adalah *Random Forest*.

Random Forest merupakan metode *ensemble learning* yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan mengurangi *overfitting* [8]. Keunggulan utama metode ini adalah kemampuannya dalam menangani data dengan dimensi tinggi serta menangkap hubungan non-linear antar variabel [9]. Hal ini menjadikan *Random Forest* sebagai metode yang potensial untuk memodelkan volatilitas pasar saham.

Dalam penelitian ini, volatilitas IHSG dihitung menggunakan pendekatan *realized volatility* yang diperoleh dari data *return* harian. Selain itu, penelitian ini juga memanfaatkan berbagai fitur teknikal yang berasal dari data historis harga saham, seperti *return*, indikator teknikal, serta *lag volatilitas*, untuk meningkatkan performa model.

Penelitian ini tidak hanya berfokus pada prediksi nilai volatilitas (regresi), tetapi juga pada klasifikasi kondisi volatilitas menjadi kategori tinggi dan rendah. Pendekatan ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai perilaku volatilitas IHSG serta membantu investor dalam mengidentifikasi kondisi pasar.

Meskipun demikian, model masih memiliki keterbatasan dalam mendeteksi lonjakan volatilitas ekstrem yang sering terjadi pada kondisi pasar yang tidak stabil [10]. Hal ini menunjukkan bahwa volatilitas pasar saham memiliki karakteristik yang kompleks dan dipengaruhi oleh faktor-faktor yang tidak sepenuhnya tercakup dalam data historis.

Berdasarkan latar belakang tersebut, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan metode prediksi volatilitas menggunakan pendekatan *machine learning*, khususnya *Random Forest*, serta memberikan wawasan bagi investor dan pelaku pasar dalam memahami dinamika volatilitas IHSG.

Dengan demikian, penelitian ini memiliki relevansi yang tinggi dalam bidang keuangan dan *data science*, serta berpotensi menjadi referensi bagi penelitian selanjutnya yang berkaitan dengan analisis volatilitas dan penerapan machine learning pada pasar modal.

2. Metode Penelitian

2.1. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan prediktif menggunakan metode *machine learning*. Tujuan utama penelitian ini adalah untuk membangun model yang mampu memprediksi dan mengklasifikasikan volatilitas Indeks Harga Saham Gabungan (IHSG) berdasarkan data historis. Pendekatan kuantitatif digunakan karena penelitian ini berfokus pada pengolahan data numerik serta pengukuran performa model menggunakan metrik statistik.

2.2. Sumber dan Jenis Data

Data adalah sekumpulan fakta atau nilai mentah yang belum diolah, yang dapat berupa angka, teks, atau simbol yang merepresentasikan suatu kondisi atau kejadian [11]. Data biasanya masih bersifat terpisah dan belum terstruktur untuk analisis [12]. Sementara itu, *dataset* adalah kumpulan data yang telah disusun secara terstruktur dalam suatu format tertentu, biasanya dalam bentuk tabel yang terdiri dari baris (observasi) dan kolom (variabel), sehingga siap digunakan untuk proses analisis atau pemodelan [13]. Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data historis harian Indeks Harga Saham Gabungan (IHSG) yang diperoleh dari platform Kaggle melalui *dataset Daily IHSG* yang dipublikasikan oleh Gareth Harrison. *Dataset* ini bersumber dari Yahoo Finance (ticker ^JKSE) dan mencakup periode panjang mulai dari 2 Januari 1995 hingga sekitar tahun 2025 dengan ribuan observasi harian. Setiap data merepresentasikan satu hari perdagangan dan terdiri dari variabel tanggal (*Date*), harga pembukaan (*Open*), harga tertinggi (*High*), harga terendah (*Low*), dan harga penutupan (*Close*). Data ini termasuk dalam kategori time series yang memiliki keterkaitan antar waktu, sehingga sangat sesuai digunakan untuk analisis prediktif, khususnya dalam pemodelan dan prediksi volatilitas pasar saham menggunakan pendekatan *machine learning* [14].

2.3. Penanganan Ketidakseimbangan Data (SMOTE)

Untuk mengatasi ketidakseimbangan kelas, digunakan metode *Synthetic Minority Over-sampling Technique (SMOTE)*. Teknik ini bertujuan untuk meningkatkan jumlah data pada kelas minoritas dengan cara membangkitkan sampel sintesis berdasarkan kedekatan antar data dalam ruang fitur [15]. Proses ini dilakukan dengan memilih titik data minoritas kemudian membuat data baru di antara titik tersebut dan tetangga terdekatnya [16].

Penggunaan *SMOTE* membantu mengurangi bias model terhadap kelas mayoritas yang umumnya mendominasi dataset [17]. Dengan distribusi data yang lebih seimbang, model diharapkan mampu meningkatkan performa dalam mendeteksi kelas minoritas, khususnya pada kasus klasifikasi volatilitas IHSG yang memiliki ketimpangan distribusi antara kondisi volatilitas tinggi dan rendah [18].

2.4. Pembagian Data

Dataset yang telah melalui tahap pra-pemrosesan selanjutnya dibagi menjadi dua bagian, yaitu data latih (training set) dan data uji (testing set) [19]. Pembagian ini dilakukan dengan proporsi tertentu, misalnya 80% untuk data latih dan 20% untuk data uji, guna memastikan bahwa model memiliki cukup data untuk proses pembelajaran sekaligus pengujian [20].

Data latih digunakan untuk membangun model dan menyesuaikan parameter algoritma, sedangkan data uji berfungsi untuk mengevaluasi performa model secara objektif terhadap data yang belum pernah digunakan sebelumnya [21]. Dengan pendekatan ini, hasil evaluasi yang diperoleh dapat mencerminkan kemampuan generalisasi model dalam memprediksi volatilitas IHSG secara lebih akurat.

2.5. Pemodelan Menggunakan Random Forest

Model klasifikasi dalam penelitian ini dibangun menggunakan algoritma *Random Forest*, yang merupakan salah satu metode *ensemble learning* [22]. Algoritma ini bekerja dengan membangun banyak pohon keputusan (*decision tree*) dari subset data dan fitur yang dipilih secara acak, kemudian menggabungkan hasil prediksi dari seluruh pohon untuk menghasilkan keputusan akhir [23].

Keunggulan *Random Forest* terletak pada kemampuannya dalam menangani data dengan kompleksitas tinggi serta mengurangi risiko *overfitting* [24]. Dengan memanfaatkan pendekatan bagging dan pemilihan fitur secara acak, model menjadi lebih stabil dan memiliki kemampuan generalisasi yang baik dalam memprediksi tingkat volatilitas IHSG berdasarkan pendekatan *realized volatility*.

2.5. Evaluasi dan Visualisasi

Evaluasi model dilakukan menggunakan beberapa metrik kinerja, seperti accuracy, precision, recall, F1-score, dan ROC-AUC [25]. Metrik-metrik ini digunakan untuk mengukur performa model dari berbagai aspek, termasuk ketepatan prediksi, kemampuan mendeteksi kelas minoritas, serta keseimbangan antara precision dan recall [26].

Selain evaluasi numerik, dilakukan pula visualisasi hasil model untuk memberikan pemahaman yang lebih komprehensif [26]. Visualisasi seperti confusion matrix digunakan untuk melihat distribusi prediksi benar dan salah, sementara kurva ROC dan precision-recall curve membantu dalam mengevaluasi kemampuan model dalam membedakan kelas volatilitas secara lebih mendalam [27].

3. Hasil dan Pembahasan

Penelitian ini menggunakan pembagian data berbasis waktu (*time series split*) yang terdiri dari periode pelatihan pada 30 Januari 1995 hingga 16 Juni 2016, periode validasi pada 17 Juni 2016 hingga 1 Februari 2021, serta periode pengujian pada 2 Februari 2021 hingga 29 September 2025. Pembagian ini bertujuan untuk menjaga sifat kronologis data sehingga model tidak mengalami *data leakage* dan dapat merepresentasikan kondisi nyata dalam prediksi volatilitas IHSG di masa depan.

Tabel 1. Hasil Evaluasi Model

Tahap	MAE	RMSE	R ²
Validasi	0.023148	0.038896	0.848852
Pengujian	0.019306	0.032344	0.807307

Berdasarkan hasil evaluasi pada data validasi, model menunjukkan kinerja yang cukup baik dengan nilai *MAE* sebesar 0.023148, *RMSE* sebesar 0.038896, dan koefisien determinasi (*R*²) sebesar 0.848852. Nilai *R*² yang mendekati 1 menunjukkan bahwa model mampu menjelaskan sebagian besar variasi data volatilitas. Selain itu, nilai error yang relatif kecil mengindikasikan bahwa prediksi model cukup dekat dengan nilai aktual, sehingga model layak untuk diuji pada data yang lebih baru.

Tabel 2. Contoh Hasil Prediksi pada Data Uji

Tanggal	Actual Volatility	Predicted Volatility	Absolute Error
2021-02-02	0.083428	0.112622	0.029194
2021-02-03	0.083165	0.102262	0.019097
2021-02-04	0.082945	0.104394	0.021449
2021-02-05	0.084029	0.098936	0.014907
2021-02-08	0.069267	0.098504	0.029237
2021-02-09	0.108918	0.087705	0.021213
2021-02-10	0.114524	0.130919	0.016396
2021-02-11	0.118403	0.129386	0.010983
2021-02-15	0.105684	0.122858	0.017173
2021-02-16	0.104080	0.125817	0.021737

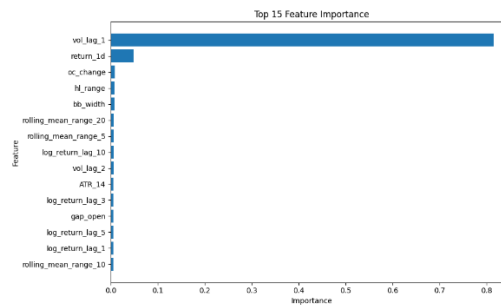
Pada tahap pengujian akhir, model menghasilkan *MAE* sebesar 0.019306, *RMSE* sebesar 0.032344, dan *R*² sebesar 0.807307. Meskipun terdapat sedikit penurunan nilai *R*² dibandingkan dengan data validasi, performa model tetap berada pada kategori baik. Hal ini menunjukkan bahwa model memiliki kemampuan generalisasi yang cukup kuat dalam memprediksi volatilitas IHSG pada data yang belum pernah dilihat sebelumnya.

Tabel 3. Top 15 Feature Importance

No	Fitur	Importance
1	vol_lag_1	0.815639
2	return_1d	0.047997
3	oc_change	0.009010
4	hl_range	0.008194
5	bb_width	0.008012
6	rolling_mean_range_20	0.007044
7	rolling_mean_range_5	0.006253
8	log_return_lag_10	0.006238
9	vol_lag_2	0.006071
10	ATR_14	0.005624
11	log_return_lag_3	0.005579
12	gap_open	0.005493
13	log_return_lag_5	0.005439
14	log_return_lag_1	0.005316
15	rolling_mean_range_10	0.005201

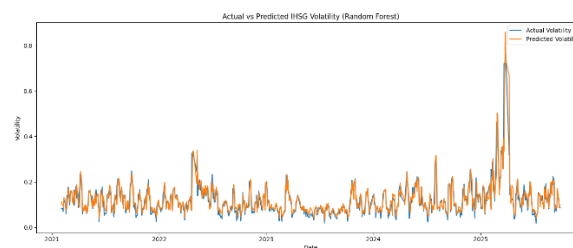
Hasil prediksi pada beberapa data awal periode pengujian menunjukkan bahwa model mampu mengikuti pola volatilitas aktual dengan cukup baik, meskipun terdapat selisih prediksi pada beberapa titik. Nilai absolute error yang relatif kecil menunjukkan bahwa deviasi antara nilai aktual dan prediksi masih dalam batas yang dapat diterima, sehingga model dapat digunakan sebagai alat bantu dalam analisis volatilitas pasar.

Analisis *feature importance* menunjukkan bahwa variabel *vol_lag_1* memiliki kontribusi paling dominan dengan nilai *importance* sebesar 0.815639. Hal ini mengindikasikan bahwa volatilitas pada periode sebelumnya sangat berpengaruh terhadap volatilitas saat ini, sejalan dengan teori *volatility clustering* dalam pasar keuangan. Selain itu, fitur lain seperti *return_1d*, *oc_change*, dan *hl_range* juga memberikan kontribusi meskipun dalam proporsi yang lebih kecil.



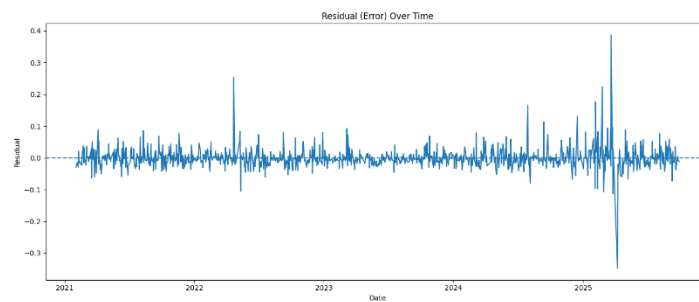
Gambar 1. Visualisasi Feature Importance

Grafik perbandingan antara volatilitas aktual dan hasil prediksi menunjukkan bahwa model *Random Forest* mampu mengikuti pola pergerakan volatilitas IHSG dengan cukup baik sepanjang periode pengujian. Garis prediksi terlihat berimpit dengan nilai aktual pada sebagian besar waktu, yang mengindikasikan bahwa model berhasil menangkap dinamika utama volatilitas pasar. Meskipun demikian, terdapat beberapa deviasi pada saat terjadi lonjakan volatilitas yang ekstrem, di mana model cenderung sedikit *overestimate* atau *underestimate*. Hal ini merupakan fenomena umum pada model berbasis pohon ketika menghadapi perubahan yang sangat tajam. Secara keseluruhan, kedekatan antara kedua kurva menunjukkan bahwa model memiliki performa yang stabil dan cukup akurat dalam merepresentasikan pola volatilitas IHSG.



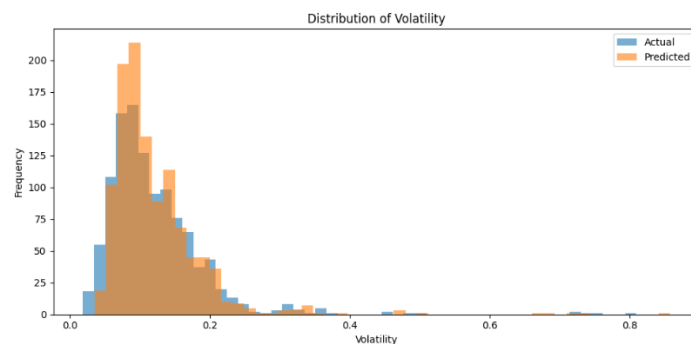
Gambar 2. Grafik Perbandingan IHSG Aktual vs IHSG Prediksi

Grafik residual (*error*) pada gambar 3 menunjukkan bahwa sebagian besar nilai residual tersebar di sekitar garis nol, yang mengindikasikan bahwa model tidak memiliki bias sistematis dalam melakukan prediksi volatilitas IHSG. Hal ini menandakan bahwa secara umum model mampu memberikan estimasi yang cukup akurat terhadap nilai aktual. Namun, terlihat adanya beberapa lonjakan residual yang cukup besar, terutama pada periode tertentu seperti sekitar tahun 2025, yang mencerminkan kesulitan model dalam menangkap perubahan volatilitas yang sangat ekstrem. Pola ini juga mengindikasikan adanya variabilitas error yang meningkat (*heteroskedastisitas*) pada kondisi pasar yang lebih bergejolak. Secara keseluruhan, distribusi residual yang relatif acak dan berpusat di sekitar nol menunjukkan bahwa model memiliki performa yang baik, meskipun masih terdapat keterbatasan dalam menangani lonjakan volatilitas yang tajam.

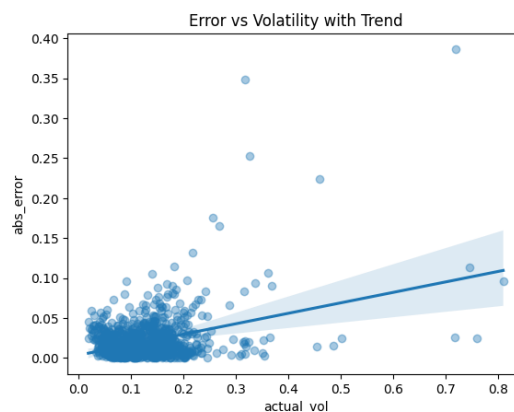


Gambar 3. Residual Overtime

Grafik distribusi volatilitas pada gambar 4 menunjukkan bahwa pola distribusi antara nilai aktual dan hasil prediksi memiliki kemiripan yang cukup tinggi, terutama pada rentang volatilitas rendah hingga menengah. Hal ini mengindikasikan bahwa model *Random Forest* mampu menangkap karakteristik utama distribusi volatilitas IHSG dengan baik. Kedua distribusi sama-sama menunjukkan kecenderungan *right-skewed*, di mana sebagian besar nilai berada pada level volatilitas rendah, sementara nilai ekstrem muncul dengan frekuensi yang lebih kecil. Namun, terlihat bahwa distribusi prediksi cenderung sedikit lebih halus dan kurang menangkap kejadian volatilitas yang sangat tinggi dibandingkan data aktual. Secara keseluruhan, kesesuaian bentuk distribusi ini menunjukkan bahwa model memiliki kemampuan yang baik dalam merepresentasikan pola statistik volatilitas, meskipun masih terdapat keterbatasan dalam menangkap ekor distribusi (*tail risk*).



Gambar 4. Distribution of Volatility

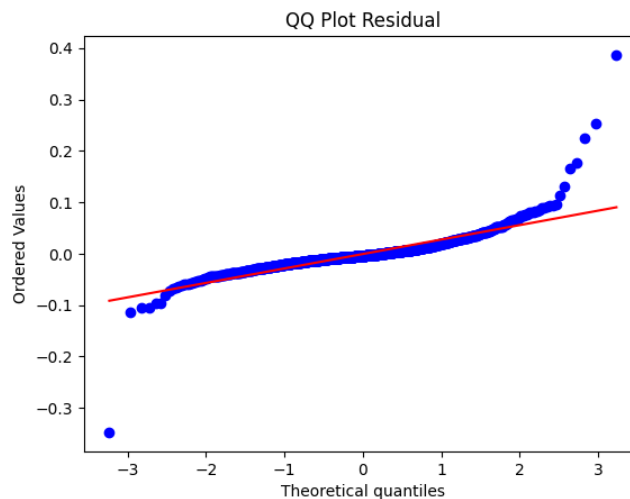


Gambar 5. Grafik hubungan antara volatilitas aktual dan absolute error

Grafik hubungan antara volatilitas aktual dan *absolute error* pada Gambar 5 menunjukkan adanya kecenderungan peningkatan *error* seiring dengan naiknya tingkat volatilitas. Hal ini terlihat dari garis tren yang menanjak, yang mengindikasikan bahwa model cenderung menghasilkan prediksi yang lebih akurat pada kondisi volatilitas rendah, namun mengalami peningkatan kesalahan pada saat volatilitas tinggi. Sebagian besar titik data terkonsentrasi pada volatilitas rendah dengan error yang kecil, menunjukkan performa model yang stabil pada kondisi pasar normal. Namun, pada nilai volatilitas yang lebih tinggi, sebaran *error* menjadi lebih luas dan terdapat beberapa outlier dengan *error* yang cukup besar. Temuan ini mengindikasikan bahwa model *Random Forest* memiliki keterbatasan

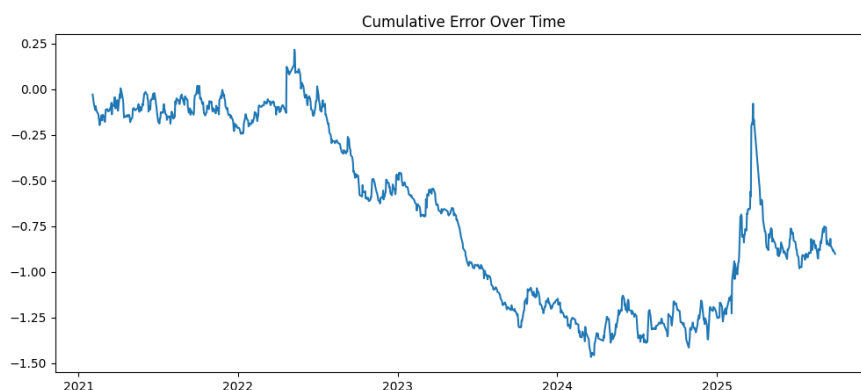
dalam menangkap dinamika ekstrem atau lonjakan volatilitas, sehingga performanya menurun pada kondisi pasar yang sangat bergejolak.

QQ plot residual menunjukkan bahwa sebagian besar titik berada di sekitar garis diagonal, yang menandakan bahwa distribusi residual mendekati distribusi normal pada bagian tengah data. Hal ini mengindikasikan bahwa model memiliki performa yang cukup baik dalam menangkap pola umum volatilitas IHSG. Namun, terlihat adanya penyimpangan yang cukup signifikan pada bagian ekor distribusi, terutama pada sisi kanan (*positif*) dan kiri (*negatif*), yang menunjukkan adanya *heavy tails* atau nilai ekstrem. Penyimpangan ini menandakan bahwa residual tidak sepenuhnya berdistribusi normal, khususnya pada kondisi volatilitas tinggi. Secara keseluruhan, meskipun asumsi normalitas tidak sepenuhnya terpenuhi, model tetap menunjukkan kinerja yang baik pada sebagian besar data, dengan keterbatasan dalam menangani kejadian ekstrem di pasar.



Gambar 6. *QQ Plot Residual*

Grafik *cumulative error* pada gambar 7 menunjukkan akumulasi kesalahan prediksi model sepanjang waktu, di mana terlihat adanya tren penurunan (negatif) yang cukup signifikan terutama sejak pertengahan periode pengujian. Hal ini mengindikasikan bahwa secara kumulatif model cenderung melakukan *underestimation* terhadap volatilitas aktual. Meskipun terdapat beberapa periode pemulihan, seperti lonjakan positif pada awal tahun 2025, tren keseluruhan masih didominasi oleh akumulasi *error* negatif. Pola ini menunjukkan bahwa meskipun error individual relatif kecil, bias kecil yang terjadi secara konsisten dapat terakumulasi menjadi deviasi yang cukup besar dalam jangka panjang. Oleh karena itu, hasil ini mengindikasikan perlunya penyempurnaan model, khususnya dalam mengurangi bias prediksi agar performa jangka panjang menjadi lebih stabil dan akurat.



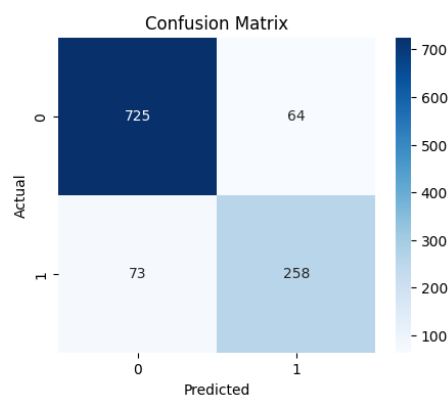
Gambar 7. *Cumulative Error*

Hasil klasifikasi volatilitas menunjukkan bahwa model *Random Forest* mampu memberikan kinerja yang cukup baik dengan tingkat akurasi sebesar 88%. Berdasarkan *classification report*, model memiliki kemampuan yang sangat baik dalam mengidentifikasi kelas mayoritas dengan nilai *precision* sebesar 0.91 dan *recall* sebesar 0.92, yang menunjukkan bahwa sebagian besar data pada kelas ini berhasil diprediksi dengan benar. Sementara itu, pada kelas minoritas, model juga menunjukkan performa yang cukup baik dengan *precision* sebesar 0.80, *recall* sebesar

0.78, dan *F1-score* sebesar 0.79, meskipun masih terdapat beberapa kesalahan klasifikasi. Hal ini juga terlihat pada gambar 9 terkait *confusion matrix*, di mana terdapat 725 prediksi benar untuk kelas mayoritas dan 258 untuk kelas minoritas, dengan sejumlah kesalahan berupa 64 *false positive* dan 73 *false negative*. Secara keseluruhan, hasil ini menunjukkan bahwa model memiliki keseimbangan yang baik dalam mengklasifikasikan kedua kelas, meskipun masih terdapat ruang perbaikan terutama dalam meningkatkan deteksi pada kelas minoritas.

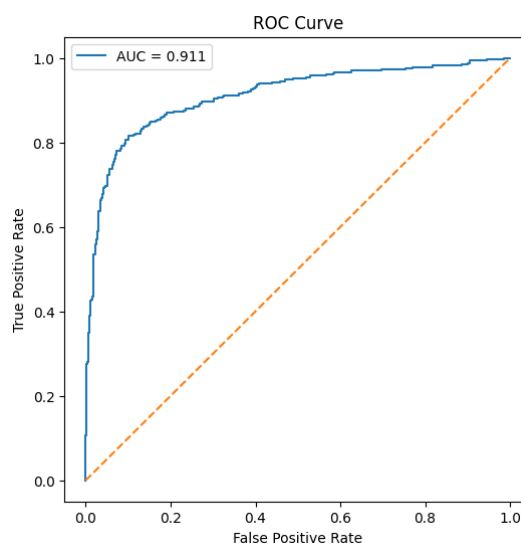
Tabel 4. *Classification Report*

Kelas	Precision	Recall	F1-Score	Support
0	0.91	0.92	0.91	789
1	0.80	0.78	0.79	331
Accuracy			0.88	1120
Macro Avg	0.85	0.85	0.85	1120
Weighted Avg	0.88	0.88	0.88	1120



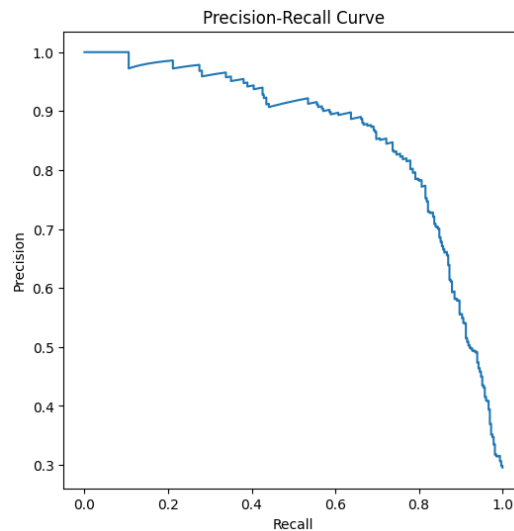
Gambar 8. Confusion Matrix

Kurva ROC pada Gambar 9 menunjukkan bahwa model klasifikasi memiliki kemampuan diskriminasi yang sangat baik dalam membedakan antara kelas volatilitas. Hal ini ditunjukkan oleh nilai *AUC* sebesar 0.911, yang berada pada kategori sangat baik dan mendekati nilai sempurna. Kurva ROC yang menjauh dari garis diagonal (*random classifier*) mengindikasikan bahwa model mampu mencapai *true positive rate* yang tinggi dengan *false positive rate* yang relatif rendah. Selain itu, bentuk kurva yang melengkung tajam ke arah kiri atas menunjukkan bahwa model efektif dalam mengidentifikasi kelas positif tanpa menghasilkan terlalu banyak kesalahan pada kelas negatif. Secara keseluruhan, hasil ini memperkuat bahwa model *Random Forest* tidak hanya akurat secara keseluruhan, tetapi juga memiliki kemampuan klasifikasi yang kuat dalam membedakan tingkat volatilitas IHSG.



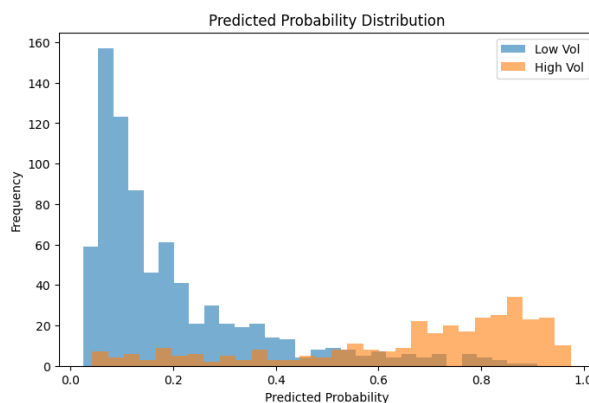
Gambar 9. ROC Curve

Kurva *Precision-Recall* pada gambar 10 menunjukkan bahwa model memiliki kinerja yang baik dalam mendeteksi kelas positif, terutama pada tingkat *recall* rendah hingga menengah. Pada bagian awal kurva, nilai *precision* berada pada tingkat yang sangat tinggi, bahkan mendekati 1, yang menandakan bahwa prediksi positif yang dihasilkan model sebagian besar benar. Seiring meningkatnya *recall*, *precision* mengalami penurunan secara bertahap, yang merupakan pola umum karena model mulai mengklasifikasikan lebih banyak data sebagai positif. Penurunan yang lebih tajam terlihat pada *recall* tinggi, yang mengindikasikan adanya peningkatan *false positive* ketika model berusaha menangkap hampir seluruh kasus positif. Secara keseluruhan, bentuk kurva ini menunjukkan adanya *trade-off* yang cukup baik antara *precision* dan *recall*, serta mengindikasikan bahwa model cukup efektif dalam menangani klasifikasi, khususnya dalam konteks data yang berpotensi tidak seimbang seperti volatilitas IHSG.



Gambar 10. Kurva Precision-Recall

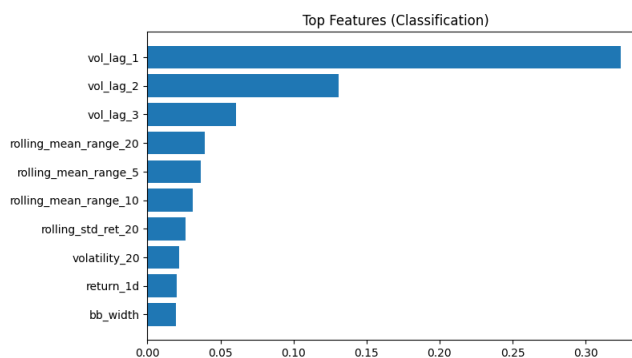
Distribusi probabilitas prediksi pada gambar 11 menunjukkan pemisahan yang cukup jelas antara kelas volatilitas rendah (*low volatility*) dan volatilitas tinggi (*high volatility*). Probabilitas untuk kelas volatilitas rendah cenderung terkonsentrasi pada nilai yang lebih kecil (dekat 0), sedangkan kelas volatilitas tinggi dominan pada nilai probabilitas yang lebih besar (mendekati 1). Hal ini mengindikasikan bahwa model memiliki tingkat kepercayaan yang baik dalam membedakan kedua kelas. Meskipun demikian, terdapat sedikit area tumpang tindih pada rentang probabilitas menengah, yang menunjukkan adanya beberapa kasus yang sulit diklasifikasikan secara pasti. Secara keseluruhan, distribusi ini menunjukkan bahwa model *Random Forest* memiliki kemampuan diskriminasi yang kuat serta menghasilkan prediksi probabilitas yang cukup terkalibrasi dalam mengklasifikasikan volatilitas IHSG.



Gambar 11. Distribusi Probabilitas Prediksi

Grafik *feature importance* pada model klasifikasi menunjukkan bahwa variabel *vol_lag_1* memiliki kontribusi paling dominan dibandingkan fitur lainnya, yang mengindikasikan bahwa volatilitas pada periode sebelumnya merupakan faktor utama dalam menentukan klasifikasi volatilitas saat ini. Fitur *vol_lag_2* dan *vol_lag_3* juga memberikan pengaruh yang cukup signifikan, memperkuat adanya pola *volatility clustering* di pasar keuangan. Selain itu, fitur berbasis statistik seperti *rolling_mean_range* dan *rolling_std_ret* turut berkontribusi dalam

menangkap dinamika pergerakan harga. Fitur lain seperti *return_1d* dan *bb_width* memiliki pengaruh yang relatif lebih kecil, namun tetap memberikan tambahan informasi bagi model. Secara keseluruhan, hasil ini menunjukkan bahwa model sangat bergantung pada informasi historis volatilitas dalam melakukan klasifikasi, dengan dukungan dari indikator teknikal lainnya untuk meningkatkan akurasi prediksi.



Gambar 12. Top Feature Classification

4. Kesimpulan dan Saran

Secara keseluruhan, pendekatan Random Forest dengan realized volatility terbukti mampu memberikan kinerja yang baik dalam prediksi dan klasifikasi volatilitas IHSG, ditunjukkan oleh nilai R^2 yang tinggi pada regresi serta akurasi klasifikasi yang mencapai 88% dengan AUC sebesar 0.911. Model mampu menangkap pola utama volatilitas, khususnya melalui kontribusi kuat variabel lag volatilitas, meskipun masih memiliki keterbatasan dalam menghadapi lonjakan volatilitas ekstrem yang tercermin dari peningkatan error pada kondisi tertentu. Oleh karena itu, penelitian selanjutnya disarankan untuk mengeksplorasi model yang lebih adaptif terhadap data deret waktu seperti LSTM atau hybrid model, serta melakukan optimasi fitur dan teknik penanganan ketidakseimbangan data yang lebih lanjut agar performa model, terutama pada kondisi volatilitas tinggi, dapat ditingkatkan.

Daftar Rujukan

- [1] R. Wulan, N. Nurpadilah, and R. Pebrian, "PENGARUH INFLASI, HARGA MINYAK DUNIA, DAN SUKU BUNGA (BI RATE) TERHADAP INDEKS HARGA SAHAM GABUNGAN (IHSG) (DATA PER BULAN PERIODE 2011-2020)," *J. Pijar*, vol. 1, no. 2 SE-, pp. 130–143, Feb. 2023, doi: 10.65096/pmb.v1i2.191.
- [2] Y. Yulianto and N. Nurmalasari, "Pengaruh Inflasi, Suku Bunga, dan Nilai Tukar terhadap Indeks Harga Saham Gabungan (IHSG) di Indonesia," *Kompeten J. Ilm. Ekon. dan Bisnis*, vol. 4, no. 4, pp. 1466–1474, Jan. 2026, doi: 10.57141/kompeten.v4i4.237.
- [3] M. Hisam, "MENAVIGASI VOLATILITAS PASAR: WAWASAN TENTANG INSTRUMEN KEUANGAN DAN STRATEGI INVESTASI," *Curr. J. Ekon. Dan Perbank. Syariah*, vol. 2, no. 2, pp. 315–328, Apr. 2024, doi: 10.32806/cecy.v2i2.248.
- [4] Kamelia Taung and Atik Budi Pariyanti, "Analisis Statistik Investor Asing dan Volatilitas Pasar Modal Indonesia," *J. Cap. Mark. Bank.*, vol. 13, no. 3, pp. 21–31, Aug. 2025, doi: 10.63607/jcemb.v13i3.21.
- [5] A. Jalil, S. Kasnelly, and I. Agustia, "Dampak Kebijakan Moneter Terhadap Stabilitas Ekonomi Di Tengah Krisis Global," *Al-Mizan J. Ekon. Syariah*, vol. 7, no. 2, pp. 105–119, Dec. 2024, doi: 10.54459/almizan.v7i11.789.
- [6] M. A. Pandu W, R. E. Saputro, P. Purwadi, and U. A. Rohmah, "Comparison of the Accuracy Levels of Naive Bayes, Random Forest, and Long Short-Term Memory (LSTM) Methods in Predicting Gold Jewelry Sales," *J. Tek. Inform.*, vol. 7, no. 1 SE-Articles, pp. 126–146, Feb. 2026, doi: 10.52436/1.jutif.2026.7.1.5139.
- [7] J. J. Hidayat and S. Hasanudin, "Prediksi Volatilitas IHSG Dengan Hybrid Model GARCH–Random Forest Berbasis Machine Learning," *J. Manaj. Inform. Teknol.*, vol. 6, no. 1 SE-Articles, pp. 130–140, 2026, doi: 10.51903/mifortekh.v6i1.1134.
- [8] A. Nugroho and D. Harini, "Teknik Random Forest untuk Meningkatkan Akurasi Data Tidak Seimbang," *JISITIK J. Sist. Inf. dan Teknol. Inf. Komput.*, vol. 2, no. 2, pp. 128–140, Jun. 2024, doi: 10.53624/jsitik.v2i2.379.
- [9] Z. Z. Hulafah Al Abrori and E. R. Subhiyakti, "Analisis Komparatif Akurasi Prediksi Kanker Payudara Menggunakan Algoritma Random Forest dan Logistic Regression," *J. Algoritma.*, vol. 22, no. 1, pp. 300–311, May 2025, doi: 10.33364/algoritma/v.22-1.2164.
- [10] E. Rafulta, F. Yanuar, D. Devianto, and Maiyastri, "Pemodelan dan Peramalan Volatilitas Memori Panjang pada Return Saham ANTM Studi Komparatif Model GARCH dan FIGARCH," *Lattice J. J. Math. Educ. Appl.*, vol. 5, no. 1 SE-Articles, pp. 75–89, Jun. 2025, doi: 10.30983/lattice.v5i1.9525.
- [11] J. T. Santoso, "ILMU DATA (Data Science)," *Penerbit Yayasan Prima Agus Tek.*, vol. 9, no. 1 SE-Judul Buku, pp. 1–265, 2023, [Online]. Available: <https://penerbit.stekom.ac.id/index.php/yayasanpat/article/view/439>
- [12] B. Mudjiyanto, "TIPE PENELITIAN EKSPLORATIF KOMUNIKASI," *J. Stud. Komun. dan Media*, vol. 22, no. 1, p. 65, Jun. 2018, doi: 10.31445/jskm.2018.220105.
- [13] A. Wibowo, "Cara Mudah Menganalisis Big Data," *Penerbit Yayasan Prima Agus Tek.*, vol. 10, no. 1 SE-Judul Buku, pp. 1–159, 2024, [Online]. Available: <https://penerbit.stekom.ac.id/index.php/yayasanpat/article/view/516>
- [14] A. Z. Kamalia, Choiriyatun Nisa Latansa, and Zaenur Rozikin, "Klasifikasi Kondisi Pasar Harga Emas ANTAM Indonesia Menggunakan Algoritma Decision Tree," *J. Komput. Teknol. Inf. Sist. Inf.*, vol. 4, no. 3, pp. 2087–2098, Jan. 2026, doi: 10.62712/juktisi.v4i3.800.
- [15] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/jair.953.

- [16] A. P. Werdana, "Pemodelan Klasifikasi Efisiensi Kalori Berbasis Data Aktivitas dan Kondisi Fisiologis Menggunakan Random Forest dan SMOTE," *J. Teknol. Inf. Digit.*, vol. 2, no. 1, pp. 54–62, [Online]. Available: <https://jurnal.ipdig.id/index.php/jtid/article/view/222>
- [17] P. Setiyadi, M. N. Prayogi, and A. Solichin, "Optimalisasi prediksi kehilangan karyawan menggunakan teknik rfe, smote, dan adaboost," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 4, pp. 2131–2145, 2024.
- [18] B. T. Pramesti, F. T. Anggraeny, and W. S. J. Saputra, "Klasifikasi Data Tidak Seimbang Menggunakan SMOTE-ENN dengan Optimasi Bayesian," *JHIP - J. Ilm. Ilmu Pendidik.*, vol. 8, no. 10, pp. 12207–12210, Oct. 2025, doi: 10.54371/jhip.v8i10.9456.
- [19] J. J. Hidayat, C. Setyowati, and A. P. Werdana, "Sentiment Analysis of Instagram User Comments related to the Inauguration of Mr. Prabowo Subianto as President of the Republic of Indonesia Using Natural Language Processing," *Int. J. Data Sci.*, vol. 6, no. 2, pp. 94–102, Dec. 2025, doi: 10.18517/ijods.6.2.94-102.2025.
- [20] J. J. Hidayat, C. Setyowati, and A. P. Werdana, "Perancangan Sistem Prediksi Penyakit pada Tanaman Padi Berbasis Image Processing Menggunakan Algoritma VGG-16 Transfer Learning dan K-Means Segmentation," *J. Pract. Comput. Sci.*, vol. 5, no. 1, pp. 1–15, May 2025, doi: 10.37366/jpcs.v5i1.5759.
- [21] J. J. Hidayat et al., "Prediksi Diabetes Menggunakan Deep Neural Network dengan Penyesuaian Hiperparameter Berbasis Bayesian Optimization," *J. Pract. Comput. Sci.*, vol. 5, no. 2, pp. 130–143, Jan. 2026, doi: 10.37366/jpcs.v5i2.6419.
- [22] Y. A. Mustofa and I. S. K. Idris, "Ensemble Approach to Sentiment Analysis of Google Play Store App Reviews," *Jambura J. Electr. Electron. Eng.*, vol. 6, no. 2, pp. 181–188, Jul. 2024, doi: 10.37905/jjee.v6i2.25184.
- [23] A. Nugroho, Wiyanto, and D. Maulana, "COMPARATIVE ANALYSIS OF CLASSIFICATION ALGORITHMS IN HANDLING IMBALANCED DATA WITH SMOTE OVERSAMPLING APPROACH," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 11, no. 2, pp. 487–495, Nov. 2025, doi: 10.33480/jitk.v11i2.6956.
- [24] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, and H. Pratiwi, "Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorar," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, pp. 563–572, Nov. 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.
- [25] M. D. I. Amin, J. J. Hidayat, C. Setyowati, E. K. Fitri, A. N. Anggraini, and A. P. Werdana, "Implementasi Model LSTM Untuk Peramalan Curah Hujan Di Bekasi Dengan Pemanfaatan Data Cuaca BMKG," *J. Teknol. Inf. Digit.*, vol. 1, no. 2, pp. 90–99, [Online]. Available: <https://jurnal.ipdig.id/index.php/jtid/article/view/200>
- [26] M. A. Rahman Wahid, A. Nugroho, and A. Halim Anshor, "Prediksi Penyakit Kanker Paru-Paru Dengan Algoritma Regresi Linier," *Bull. Inf. Technol.*, vol. 4, no. 1, pp. 63–74, Mar. 2023, doi: 10.47065/bit.v4i1.501.
- [27] A. M. Rifai, S. Raharjo, E. Utami, and D. Ariatmanto, "Analysis for diagnosis of pneumonia symptoms using chest X-ray based on MobileNetV2 models with image enhancement using white balance and contrast limited adaptive histogram equalization (CLAHE)," *Biomed. Signal Process. Control*, vol. 90, p. 105857, Apr. 2024, doi: 10.1016/j.bspc.2023.105857.